

Weekly Report

29nd April, 2018

Done:

1. Reimplement the timeline view, reconstruct its components. Modify and integrate a new uncertainty data layer into the system.
2. 3rd week of Hyperparameter tuning (with assignments uses TensorFlow) and obtain the certificate. Finish the homework of 1st week of CNN.
3. Finish the review comment of vast paper.
4. Discuss the ideas of graph DOI. First, if we apply this idea: (1) for the API (a priori interest), we could use things like: fires that have large areas, long times, or are more recent could have a higher API and/or those that have higher node degrees (2) for the UI (user interest), we could get some information by talking to the emergency management people when we have the meeting with them (3) and the idea of using their search terms to help determine the UI is also good: if they are searching for things in a particular geographical area, that area would get a higher UI.

Second, it's still too vague, and it will take a lot of time to implement the two basic prototypes. Now that we don't know the exact requirement of target users we should do something more feasible like first try some dimension reduction or clustering methods. There is a VIS17 paper: Towards a Systematic Combination of Dimension Reduction and Clustering in Visual Analytics. We may start there. Also, I am trying to learn more about embeddings.

Third, Max is also trying to do something like graph evolution of fire hotspots.

To Do:

1. Trying to store the weather data into our DB for easier analysis and modeling. (with the help of Alex); the polygon selection is a functionality to be added but difficult..
2. Read more papers as mentioned above.

Paper reading:

AAAI18 *Anchors: High-Precision Model-Agnostic Explanations* 这是 LIME 作者的最新作品, 是对模型局部规律的更精确描述. 它建立一套规则集合, 使得在预测样本的邻域(扰动的邻域)上有很高的置信度, 这样的规则集合就是 Anchors. 他比之前 LIME(局部的线性模型)复杂或者说精细. 但是其实验表明在文本上效果比较好. 其代码同样有了 github repo.

AAAI18 *Beyond Sparsity: Tree Regularization of Deep Models for Interpretability* 这是另一个模型解释性的文章, 其对正则化项做了一些改进. 它区别于别的方法, 包括上面这类文章在于, 上面一类文章经常是在一个扰动邻域里做文章, 而本文是致力于在训练时让模型变得跟可解释. 由于其 RNN 一些机制还比较复杂, 差不多下周开始学习 coursera 上的序列模型, 后面会再来回顾.

TVC18 *TOPKUBE: A Rank-Aware Data Cube for Real-Time Exploration of Spatiotemporal Data* 从 data cube 到 nanocubes 到 topkube 一层层进步. 本文关注点在于查询 k 个最相关

结果, 并让其效率达到可以交互的要求.